PNAS RESEARCH ARTICLE | GENETICS

The evolution of the human DNA replication timing program

Alexa N. Bracci[a], Anissa Dallmann[a], Qiliang Ding[a], Melissa J. Hubisz[b], Madison Caballero[a], and Amnon Koren[a,1]

Edited by Marcus Feldman, Stanford University, Stanford, CA; received August 15, 2022; accepted January 23, 2023

DNA is replicated according to a defined spatiotemporal program that is linked to both gene regulation and genome stability. The evolutionary forces that have shaped replication timing programs in eukaryotic species are largely unknown. Here, we studied the molecular causes and consequences of replication timing evolution across 94 humans, 95 chimpanzees, and 23 rhesus macaques. Replication timing differences recapitulated the species' phylogenetic tree, suggesting continuous evolution of the DNA replication timing program in primates. Hundreds of genomic regions had significant replication timing variation between humans and chimpanzees, of which 66 showed advances in replication origin firing in humans, while 57 were delayed. Genes overlapping these regions displayed correlated changes in expression levels and chromatin structure. Many human–chimpanzee variants also exhibited interindividual replication timing variation, pointing to ongoing evolution of replication timing at these loci. Association of replication timing variation with genetic variation revealed that DNA sequence evolution can explain replication timing variation between species. Taken together, DNA replication timing shows substantial and ongoing evolution in the human lineage that is driven by sequence alterations and could impact regulatory evolution at specific genomic sites.

comparative genomics | replication timing | human evolution

Significance

Genomic DNA replicates during the S phase of the cell cycle in a consistent order known as DNA replication timing. Replication timing is associated with gene regulation and influences mutation rates; however, the impacts of replication timing on human evolution are largely unknown. Here, we generated replication timing for humans, chimpanzees, and rhesus macaques and characterized variation within and between species. We identified hundreds of evolutionary replication timing variants that corresponded with gene regulation evolution and incurred elevated mutation rates. Linking genetic variation to replication timing variation facilitated the identification of sequence changes driving replication timing evolution. This study sheds light on the genetic causes and regulatory consequences of human replication timing evolution, which could impact human-specific traits.

Understanding of human-specific phenotypes and their evolution has primarily focused on the comparison of genes or sequence elements, and their regulation, between humans and closely related species (1, 2). Humans and chimpanzees are approximately 99% identical at the single-nucleotide level yet have undergone extensive phenotypic divergence (3). This has increasingly been attributed to regulatory evolution, including gene expression, which has been associated with brain, skeletal, and other phenotypes (4–7). An understudied form of genome regulation, with a potential impact on regulatory and sequence evolution, is the spatiotemporal program of DNA replication.

Genome replication is accomplished by replication origins that fire at different times during the S phase, resulting in a defined pattern of DNA replication timing. Although some origin usage appears to be influenced by stochastic (i.e., probabilistic) factors, single-cell analysis has shown that replication timing and origin firing is predominantly deterministic and predictable (8). Early DNA replication is associated with high gene density, open chromatin, and active transcription (9), while later-replicating regions typically exhibit higher frequencies of single-nucleotide mutations and polymorphisms (10–12). Replication timing thus bridges between genome regulation and evolution. As a corollary, understanding the evolution of replication timing can reveal the selective forces that have shaped particular replication programs, inform mechanisms of replication timing regulation, and uncover impacts of replication timing on sequence, molecular, and phenotypic evolution.

Only a handful of studies have compared replication timing across species. Studies in yeast suggested that replication origins dynamically gain and lose activity during evolution (13) and that conserved early replication, in particular of histone genes, is required for high gene expression levels (14). In contrast, replication timing has been shown to be highly conserved between corresponding cell types of humans and mice despite extensive genome rearrangements (15, 16), while a more recent study suggested the presence of both conserved and species-specific replication timing regions among five primate species (17). Importantly, previous studies have been underpowered to identify the genetic changes that drive replication timing evolution or its potential impacts on regulatory and sequence evolution.

Here, we address the causes and consequences of replication timing evolution by profiling a large number of humans, chimpanzees, and rhesus macaques. We find that replication timing has continuously evolved across these species at hundreds of locations. Comparison to intraspecies variation and sequence polymorphisms within species and divergence between species revealed the genetic basis of a subset of replication origins that

Deposited as a preprint on bioRxiv: doi: https://doi.org/10.1101/2022.08.09.503365.

Author contributions: A.K. designed research; A.N.B. performed research; A.N.B., A.D., Q.D., M.J.H., and M.C. analyzed data; and A.N.B. and A.K. wrote the paper.

The authors declare no competing interest.

This article is a PNAS Direct Submission.

[1]To whom correspondence may be addressed. Email: koren@cornell.edu.

This article contains supporting information online at https://www.pnas.org/lookup/suppl/doi:10.1073/pnas.2213896120/-/DCSupplemental.

Published February 27, 2023.

have gained or lost activity during evolution. On the other hand, analysis of gene expression and chromatin structure suggests a complex relationship between the evolution of replication timing and gene regulation. Overall, this study advances our knowledge on how replication timing evolves, the association of replication timing with genome regulation and transcription, and the determinants of replication timing evolution.

Results

High-Resolution DNA Replication Timing Profiles across Humans, Chimpanzees, and Rhesus Macaques. To study the evolution of DNA replication timing across primates, we sequenced the genomes of 90 chimpanzee lymphoblastoid cell lines (LCLs), 23 rhesus macaque LCLs, and seven chimpanzee induced pluripotent stem cell lines (iPSCs) along with 88 human LCLs and eight human iPSCs. We aligned each species' sequencing reads to its own reference genome and inferred DNA replication timing from read depth fluctuations across chromosomes (18–20). Specifically, DNA replication leads to a transient increase in relative copy number, which depends on the replication timing of a given genomic region: Earlier-replicating loci will have double the copy number for longer times during the S phase than later-replicating loci. This differential copy number change can be observed in whole-genome sequence data from proliferating cell samples and utilized to extract whole-genome replication timing profiles. Practically, this is achieved by calculating sequencing read depth in uniquely alignable windows, filtering out copy number variants (CNVs) and other copy number outlier regions, and then smoothing the resulting profiles. Normalization to an autosome-wide data mean of zero and SD of one converts the profiles to a standard scale that can be interpreted in terms of progression along the S phase and compared among samples (*Methods*). Our method of inferring replication timing from whole-genome sequence data was particularly suited to the evolutionary comparison of replication timing as chimpanzee material is scarcely available for the experimental manipulations required by other approaches [e.g. Repli-seq; (21)]. One chimpanzee LCL, one chimpanzee iPSC, and two human iPSCs were filtered due to low data quality. Read depth fluctuations showed long-range continuity along chromosomes (autocorrelation) consistent with DNA replication, and LCL data resolution was further improved using principal component (PC) regression (20) (*SI Appendix*, Fig. S1 *A–E*). The resulting replication timing profiles were highly consistent across samples within each species (human LCLs r = 0.94 to 0.99, chimpanzee LCLs r = 0.84 to 1, rhesus LCLs r = 0.97 to 1, human iPSCs r = 0.91 to 0.97, and chimpanzee iPSCs r = 0.96 to 0.97) (Fig. 1 *A–E* and *SI Appendix*, Fig. S2 *A–D*).

We further validated the replication timing profiles in three ways. First, we measured replication timing by sorting and sequencing G1 and S phase cells of select samples (11), which provided replication timing profiles highly correlated to those generated without cell sorting (human LCL mean r = 0.97, chimpanzee iPSC mean r = 0.87, and rhesus LCL mean r = 0.95) (Fig. 1 *A* and *C* and *SI Appendix*, Fig. S2*B*). Second, we compared the

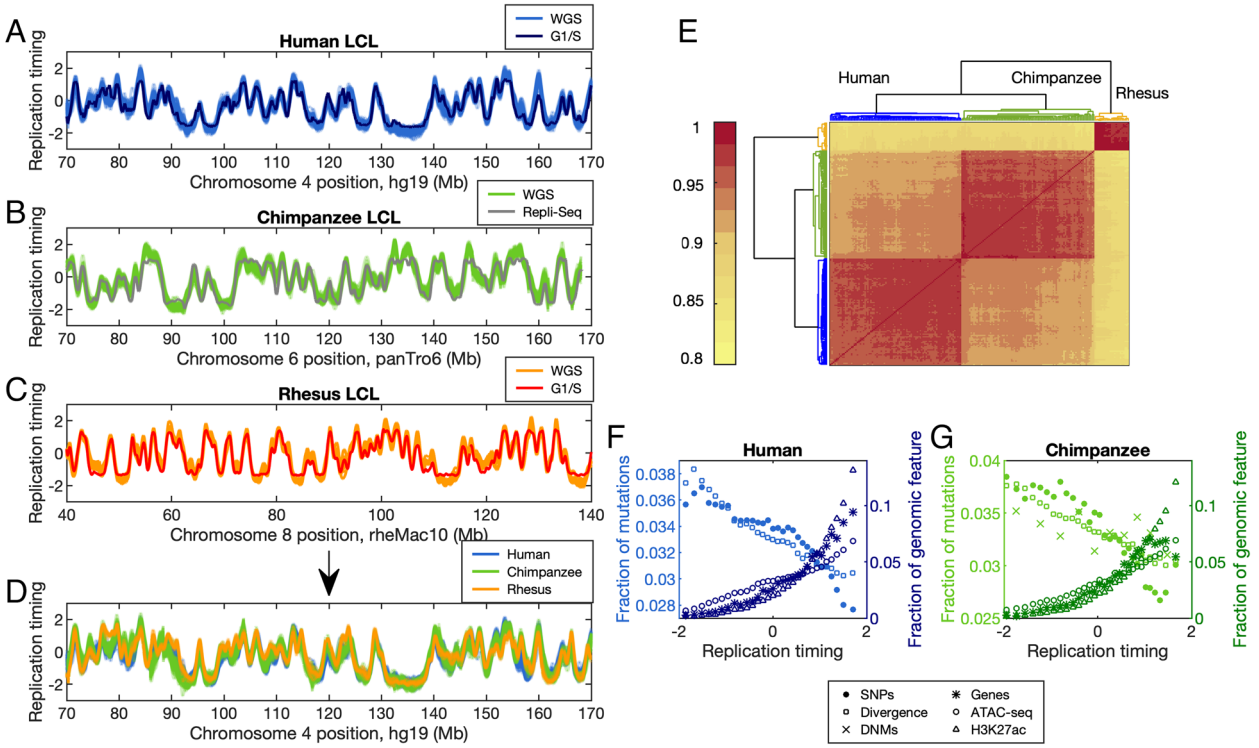

Fig. 1. Replication timing evolution in primate species. (*A–C*) Replication timing was inferred from read depth fluctuations in whole-genome sequencing (WGS) data of human, chimpanzee, and rhesus macaque LCLs. Units are SD from an autosome-wide mean of 0. Shown for comparison are a consensus G1/S profile for human LCLs (*A*) (11), a G1/S rhesus macaque LCL profile (generated in this study; *C*), and a chimpanzee LCL replication profile generated using Repli-Seq (*B*) (17). (*D*) Human, chimpanzee, and rhesus macaque replication timing profiles, plotted on human genomic coordinates (hg19; see also *SI Appendix*, Fig. S3), show conservation of the replication timing program. (*E*) Hierarchical clustering of human, chimpanzee, and rhesus macaque LCL Pearson correlation values. Replication timing is highly consistent within species and, while is largely conserved among species, exhibits significant interspecies variation that corresponds to the evolutionary divergence of primates. (*F* and *G*) SNPs, human–chimpanzee divergent sites, and de novo mutations (DNMs) are enriched at late-replicating DNA, while protein-coding genes and marks of accessible chromatin (ATAC-seq peaks and H3K27ac ChIP-seq peaks) are enriched at early-replicating DNA. Fraction of human (*F*) or chimpanzee (*G*) genomic features in 30 replication timing bins per species. DNM rate was calculated in 10 replication timing bins (fraction is 3×).

chimpanzee LCL samples to a previously published chimpanzee replication timing profile generated using Repli-seq (17) (mean r = 0.90) (Fig. 1*B*). Finally, as we showed previously in humans (22), replication of the X chromosome was delayed and less structured in chimpanzee LCL females compared to males (*SI Appendix*, Fig. S1 *G–J*). Together, these results demonstrate that the replication timing profiles of all three species are of high quality and reproducibility.

Next, we compared genome-wide replication timing to human–chimpanzee divergence and single-nucleotide polymorphisms (human dbSNP 153 common, n = 9,585,612) and found that all were enriched at late-replicating genomic regions in both human LCLs and iPSCs (Fig. 1*F* and *SI Appendix*, Fig. S2 *H*). On the other hand, gene density, gene expression levels (23–26), ATAC-seq (27), and H3K27ac ChIP-seq (25) were all enriched at, or correlated with, early-replicating genomic regions in humans (Fig. 1*F* and *SI Appendix*, Fig. S2*J*). As further biological validation, these genome-wide trends were also replicated in chimpanzee LCLs and iPSCs (chimpanzee dbSNP, n = 1,468,866) (23–27) (Fig. 1*G* and *SI Appendix*, Fig. S2 *I* and *J*). Additionally, somatic cell line mutations (identified as de novo mutations (DNMs) in chimpanzee trios; see *Methods*) were enriched at late replicating genomic regions in chimpanzee LCLs (Fig. 1*G*). Overall, this supports the conservation of genomic features associated with replication timing across cell types and species.

Substantial Variation in Replication Timing between Species. To compare replication timing profiles between species, we converted the chimpanzee and rhesus macaque replication timing data to human genome coordinates (see *Methods*); these conversions had a minimal effect on the structures of the replication profiles (*SI Appendix*, Fig. S3 *A* and *B*). We found that replication timing was highly conserved across species (Fig. 1 *D* and *E* and *SI Appendix*, Fig. S2 *C* and *D*) and that replication timing variation was greatest between cell types (LCL and iPSC; mean r = 0.69) (*SI Appendix*, Fig. S2 *E–G*). Nonetheless, there were also clear interspecies differences within the same cell type. Hierarchical clustering of replication timing values across samples recapitulated the phylogenetic tree for these three species (Fig. 1*E* and *SI Appendix*, Fig. S2*D*), suggesting that replication timing has evolved continuously across the primate lineage, primarily in a cell-type–specific manner.

We systematically searched for specific differences in replication timing between humans and chimpanzees, separately for LCLs and iPSCs, using sliding ANOVA tests with a Bonferroni corrected *P*-value threshold of 8.7×10^{-7} (see *Methods*). For the X chromosome, males and females were considered separately. We identified 858 autosomal regions where human and chimpanzee LCLs significantly differed in replication timing. These variants were at least 204 kb long and encompassed replication timing differences as small as 0.2 SD units. These regions covered 1.1 Mb on average and cumulatively spanned 980 Mb (36.6% of the analyzable autosomes). Similarly, we identified 47 variant regions on the X chromosome in females and 39 in males (1.4 and 1.2 Mb on average, spanning a total of 64 (42.9% of the analyzable X chromosome) and 45 Mb (29.9%) in females and males, respectively). In iPSCs, we identified 704 autosomal variant regions covering 1.1 Mb on average and cumulatively spanning 797 Mb (29.8% of the autosomes), likely less than in LCLs due to the more limited sample size.

A majority of the human–chimpanzee variant regions occurred at peaks in the replication timing profiles (LCL: 620/944, 65.7%; iPSC: 476/704, 67.6%), suggesting that a major source of replication timing variation is changes in replication origin (or origin cluster) activity. Extending from this observation, and under the assumption that changes in origin activity are the most likely explanation for the large replication timing variants that we observed, we designated the center of the peak as the most likely source of replication timing variation within each region (see *Methods*). This was only applied to variant regions that contained replication timing peaks, while 134 LCL variants with more than one peak were separated into several independent variant regions. Overall, we called 731 LCL and 557 iPSC replication timing variants that each contained one putative source site (Fig. 2 and *SI Appendix*, Figs. S4 and S5 and Dataset S1) and utilized them for downstream analyses. These variants covered on average 1.2 Mb in LCLs and 1.1 Mb in iPSCs and had an average magnitude of replication timing difference between humans and chimpanzees of 0.4 SDs in LCLs (*SI Appendix*, Fig. S4 *B* and *C*) and 0.5 in iPSCs. The majority of these variant regions were earlier replicating in humans compared to chimpanzees in both LCLs (57.0%; binomial test $P = 1.2 \times 10^{-4}$; Fig. 2*E* and *SI Appendix*, Fig. S4*B*) and iPSCs (53.9%; $P = 0.08$). Consistent with these variants containing replication profile peaks, the distribution of replication timing at variants was skewed toward early replication in both humans and chimpanzees (*SI Appendix*, Fig. S4*A*). The fraction of replication timing peaks that varied between humans and chimpanzees—32.5%—was comparable to the fraction of the genome with replication timing variation; thus, the widespread evolution of replication timing is not an inflated estimate due to the broad effect of individual replication origins.

In order to infer whether each replication timing change occurred in the human or chimpanzee lineage, we compared the average LCL replication timing profile for each species to the average profile of rhesus macaque as an outgroup. Specifically, we calculated the pairwise Euclidean distance of replication timing values between each pair of species within each variant region source site (*Methods*). Human-specific replication timing changes were defined as regions in which chimpanzees and rhesus macaques were closer to each other than either were to humans, and chimpanzee-specific changes were similarly defined as cases in which humans and rhesus macaques were the most similar (*Methods*). Regions with significantly different replication timing among all three species were considered unresolved for evolutionary direction. In total, we resolved 233 replication timing variant regions (of the 731 LCL variants containing a replication timing peak), of which 123 and 110 were changes in the human and chimpanzee lineages, respectively [similar number of changes in each lineage expected based on the molecular evolutionary clock; $\chi^2 = 0.73$, df = 1, and $P = 0.39$ (28)]. Of these, we inferred 66 to be human advances and 57 to be human delays, while another 31 and 79 were inferred to be replication timing advances or delays, respectively, in chimpanzees (Fig. 2 and *SI Appendix*, Fig. S4 *D* and *E*). Of the 66 human advances, 55 represented earlier activation of a shared origin, while another 11 regions appeared to represent de novo evolutionary emergence of novel replication origins in the human lineage. Similarly, 55 replication origins appeared to have been delayed in their firing time in humans compared to chimpanzees, with evidence for two replication origins being entirely lost in some humans. Notably, none of the 13 human origin gains and losses appeared to be fixed changes but were rather polymorphic among humans (*SI Appendix*, Fig. S6). This suggests that these origins have been recently gained or lost and are subject to ongoing evolution in the human lineage. In comparison, we identified seven and 56 origins that have been putatively gained or lost, respectively, in virtually all human and chimpanzee samples compared to macaques (allowing up to 5% technical variation of samples). Thus, on a broader evolutionary timescale, we see

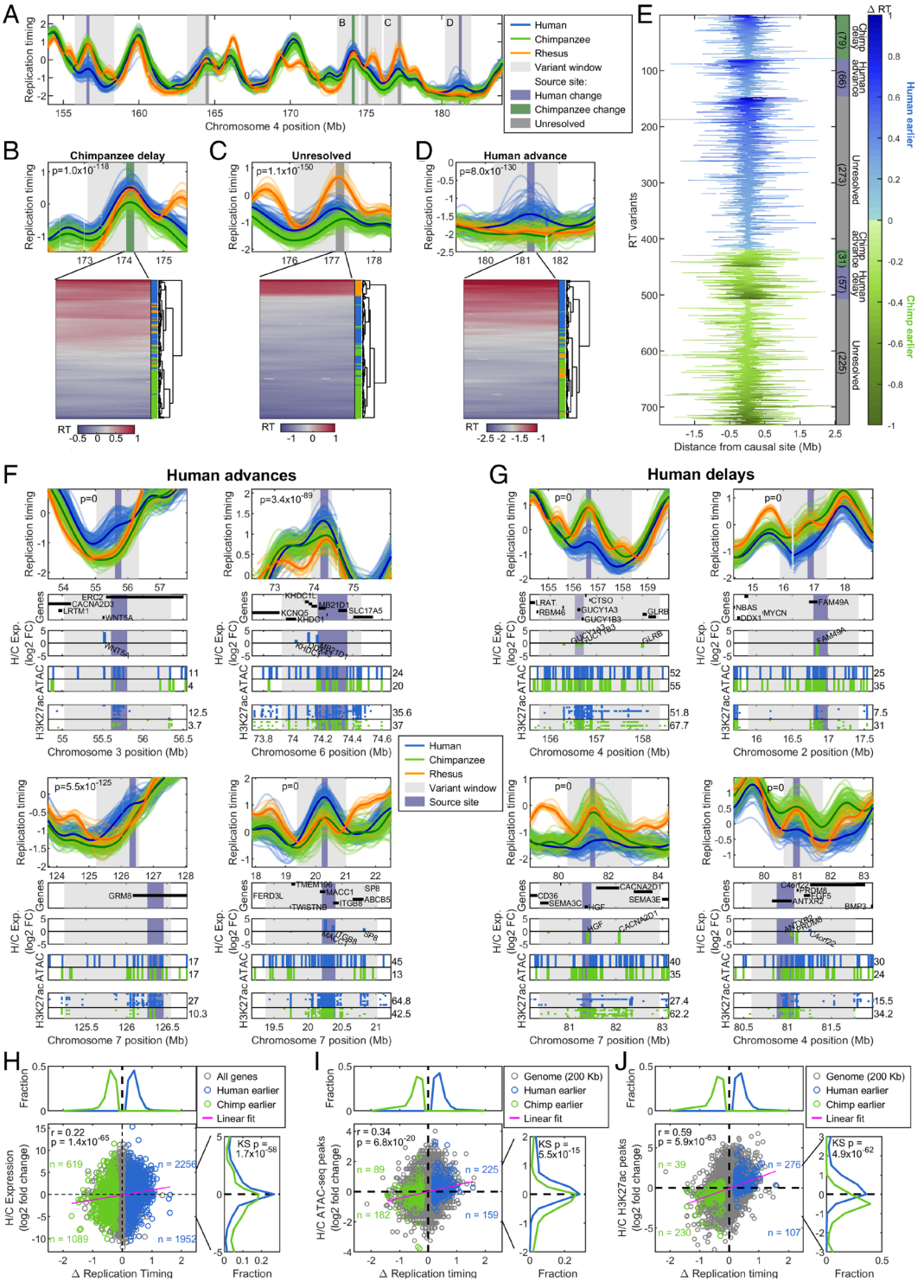

Fig. 2. Replication timing evolution and its covariation with gene expression and chromatin accessibility. (*A*) Replication timing profiles for a region of chromosome 4 for humans (n = 88, blue), chimpanzees (n = 89, green), and rhesus macaques (n = 23, orange) along with identified human–chimpanzee replication timing variant regions (light gray) and their called source sites. Source site color indicates the lineage in which replication timing was inferred to have evolved. (*B–D*) Three example regions indicated in (*A*) are shown at greater resolution. Replication timing of each sample within the source site shown as heat maps; dendrograms: hierarchical clustering of sample replication timing similarity. The clustering demonstrates the separation of the majority (or all) of human from chimpanzee samples and the clustering of one (or none) of them to rhesus macaque replication timing. *P* values: significance (ANOVA) of human–chimpanzee differences within the variant region. (*E*) Mean difference in replication timing (ΔRT) between humans and chimpanzees across each replication timing variant centered at the source sites. Variants are sorted by being earlier in humans or chimpanzees, then by the species in which the change was inferred to have happened, and last by the magnitude of interspecies replication timing difference (ΔRT). (*F* and *G*) Examples of replication timing advances (*F*) and delays (*G*) inferred to have occurred in the human lineage. Genes, expression, ATAC-seq peaks, and H3K27ac ChIP-seq data shown beneath the replication timing profiles for each variant and flanking 200 kb. Numbers next to ATAC-seq track: the number of human and chimpanzee ATAC-seq peaks within the variant window. Numbers next to H3K27ac track: average number of human and chimpanzee H3K27ac ChIP-seq peaks within the variant window. Some gene names were removed from the Genes track for readability. (*H*) Differences between human and chimpanzee LCL replication timing compared to differences in gene expression for all genes and genes within replication timing variants with either earlier replication timing in humans (blue) or in chimpanzees (green). Correlation coefficient (r) and *P* value indicated for variant regions. Number of genes in each quadrant further demonstrates the correlation between gene expression and replication timing variation. *Top* and *Right* histograms: Distributions of replication timing and gene expression differences, respectively, within replication timing variant regions. (*I* and *J*) As in *H*, using ATAC-seq (*I*) or H3K27ac ChIP-seq (*J*) data.

compelling evidence of more substantial restructuring of the replication program in primates.

Of 731 human–chimpanzee LCL replication timing variants, 47 (6.4%) were shared in iPSCs; of these, 30 had a similar shape of replication timing profiles (by correlation; see *Methods*) between cell types of the same species (*SI Appendix*, Fig. S5*E*). Thus, although variants shared across cell types have greater potential to impact species-specific phenotypic differences, most human–chimpanzee replication timing differences are cell-type specific.

Association of Replication Timing Variation with Gene Evolution. iPSC variant regions spanned 2,801 protein-coding genes (mean of five genes per variant), while LCL replication timing variants spanned a total of 6,156 protein-coding genes (mean of 8.5 genes per variant). iPSC replication timing variant regions were enriched for genes involved in immunity and development (Dataset S2), some of which are known to be under adaptive evolution in humans (e.g., AKAP11, GLB1L2, SYBU, CD59, PYHIN1, PYDC2, SIGLEC9/L1, ADAM2, OVGP1, SEMG1, SEMG2, and ANG) (29, 30). Similarly, several genes inferred to be under positive selection in humans fell into LCL replication timing variant regions. These genes included several with roles in cell cycle progression (TLE6; *SI Appendix*, Fig. S7*H*), Wnt signaling (TLE4), and sperm motility (CATSPER1, SEMG1, and SEMG2), and several associated with human diseases or conditions including Usher syndrome (USHBP1; *SI Appendix*, Fig. S7*I*), glaucoma (RMDN2; *SI Appendix*, Fig. S7*J*), intellectual disability (KPTN), and microcephaly (ASPM; *SI Appendix*, Fig. S7*K*). One notable LCL variant region spanned the APOBEC cluster that includes APOBEC3A, APOBEC3B, APOBEC3C, APOBEC3D, APOBEC3F, APOBEC3G, and APOBEC3H; these genes play a role in antiviral activity, and most have been under positive selection in primates (31) (*SI Appendix*, Fig. S7*G*). APOBEC genes replicated earlier in humans, and most fell within the source site of replication timing variation.

We identified 877 protein-coding genes with variable replication timing between humans and chimpanzees in both LCLs and iPSCs. One notable gene under positive selection in humans, PYHIN1 (IFIX), replicated later in humans compared to chimpanzees (and macaques; *SI Appendix*, Fig. S7*L*). This gene is a known tumor suppressor, downregulation of which is associated with breast cancer (32).

As a complementary analysis, we examined replication timing evolution at various genomic elements previously described to undergo atypical rates of evolution. These included human–chimpanzee divergent sites, more specifically, human accelerated regions (HARs), which are conserved in mammals yet have undergone many sequence changes in humans (33), and regions identified as under ancient positive selection in humans [selective sweeps (34)]. Conversely, we analyzed regions under evolutionary constraint: loss-of-function–intolerant genes (gnomAD) (35) and ultraconserved elements (UCEs) that are completely conserved in sequence across the human, mouse, and rat (36). Sites of sequence divergence (Fig. 1 *F* and *G*) and HARs (although not sites of selective sweeps; *SI Appendix*, Fig. S7 *C* and *D*) were biased to late replication, while UCEs and loss-of-function–intolerant genes replicated earlier than expected (compared to genes in general in the latter case; *SI Appendix*, Fig. S7 *A* and *B*). More significantly, we also found that divergent sites and HARs (but not regions under ancient positive selection in humans) were enriched in replication variant regions (focusing on variants in iPSCs—the cell type better reflecting the germ line; *SI Appendix*, Fig. S7 *C*, *D*, and *F*). On the other hand, loss-of-function–intolerant genes, as well as all protein-coding genes, were found to be significantly depleted in iPSC replication timing variant regions (*SI Appendix*, Fig. S7 *B* and *E*). Taken together, these results suggest that replication timing alterations are unfavorable at conserved regions, possibly because they have an impact on genome function. Conversely, sequence divergence appears to be associated with replication timing differences between species.

Complex Association between DNA Replication Timing and Gene Regulation. Since replication timing is correlated with genome regulation (e.g., gene expression and chromatin accessibility; Fig. 1), we tested whether replication timing variation was itself correlated with differences in gene expression or chromatin accessibility. Indeed, replication timing differences were positively correlated with gene expression variation (LCL: r = 0.22; iPSC: r = 0.25), and most replication timing variants (LCLs: 407/731, 56%, and z test $P = 7.0 \times 10^{-4}$; iPSCs: 312/557, 56%, and z test $P = 9.2 \times 10^{-4}$) contained predominantly genes with interspecies gene expression variation that corresponded to the direction of replication timing variation (i.e., earlier replication associated with elevated gene expression and later replication with reduced gene expression) (Fig. 2*H* and *SI Appendix*, Fig. S5*A*). Similarly, we observed a positive correlation between replication timing variation and chromatin accessibility assessed using ATAC-seq (LCL: r = 0.35) and the histone modifications H3K27ac (LCL: r = 0.59; iPSC: 0.53), H3K4me1 (LCL: r = 0.44), and H3K4me3 (LCL: 0.51): The earlier-replicating species had a relatively higher density of the open chromatin marks compared to the later-replicating species (Fig. 2 *I* and *J* and *SI Appendix*, Figs. S4 *F* and *G* and S5*B*). In contrast, density of the repressive chromatin mark H3K27me3 was not significantly correlated with replication timing variation (LCL: 0.05; iPSC: −0.09; *SI Appendix*, Figs. S4*H* and S5*C*). Overall, 90% of autosomal human–chimpanzee replication timing variant regions had concordant changes in replication timing and either gene expression or chromatin structure (based on H3K27ac, the histone mark most correlated to replication timing), and 52% (343/656) had concordant changes in all three. Thus, beyond the genome-wide correlations, we see specific covariation of replication timing, chromatin structure, and gene expression levels. Nonetheless, in many other instances, replication timing and gene expression or chromatin structure variations were decorrelated (e.g., Fig. 2 *H–J*; *Upper Left* and *Lower Right* quadrants and uncorrelated data points), underscoring the complexity of these fundamentally separate epigenetic processes.

To get a better understanding of the relationships between DNA replication timing and chromatin, we analyzed their spatial covariation. Namely, we asked whether chromatin differences between species better align with replication origin sites specifically or rather with entire "replicons" (the regions between an origin and its two neighboring termini). While not providing definitive proof of causality, spatial analysis suggests likely direction of effects. In some variant regions, chromatin structure differed between species primarily at the source site of replication timing variation (Fig. 2 *F*, *Top Left* and Fig. 2 *G*, *Top Right*), suggesting that chromatin structure could be a determinant of the observed replication timing variation. In contrast, in other instances, differential chromatin structure/accessibility was present across the entire variant region (e.g., Fig. 2 *G*, *Top Left* and *Bottom Left*), suggesting that instead replication timing could be exerting long-range effects on chromatin structure. Similarly, there was no consistent spatial relationship between replication timing and gene expression variation; in some cases, gene expression varied concordantly primarily at the source site of replication timing variation (Fig. 2 *F*, *Bottom Right* and Fig. 2 *G*, *Top Right*), while in

other cases, concordant replication timing–gene expression variation extended across the entire variant (Fig. 2 *F*, *Top Right* and Fig. 2 *G*, *Bottom Left*). The evidence for each of these patterns across numerous genomic regions suggests that the interaction between replication timing and gene expression regulation is complex and locus specific. As an extension, gene expression variation was not generally higher for genes in replication timing variant regions compared to nonvariant regions (mean log FC of genes in variants = 1.3 and nonvariants = 1.4) together indicating that gene expression and replication timing variation, while often linked, are neither sufficient nor necessary drivers of one another.

Genetic Basis of Replication Timing Evolution. The differences between species described above are suggestive of past and/or ongoing evolution of replication timing. As an extension of this observation, ongoing evolution is expected to manifest as interindividual variation within a given species. Indeed, we have previously shown that replication timing varies among humans at hundreds of genomic locations (18, 20). Consistently, in the current LCL sample set, we identified 185 human and 195 chimpanzee genomic regions with significant variation among individuals (*Methods* and Fig. 3). Of those, 73 regions varied among individuals in both species, significantly more than expected by chance (18 expected; z test $P = 6.5 \times 10^{-40}$) (Fig. 3 *C*, *F*, and *G*), while 112 regions were variable only among humans and 122 only among chimpanzees (Fig. 3 *D*, *E*, and *G*). More than half of the intraspecies variants were also identified as interspecies variants (Fig. 3 *D*–*G*), including variants that were shared across species (40/73; 22 expected; z test $P = 7.8 \times 10^{-6}$) or those that were species specific (63/112 humans; 34 expected; z test $P = 4.8 \times 10^{-9}$; 57/122 chimpanzees; 34 expected; z test $P = 1.1 \times 10^{-6}$). When directly testing human–chimpanzee variants for within-species variation, 239 variants were also polymorphic in at least one of the species. Notably, 20 resolved human-evolved variants were also variable among humans, suggesting ongoing evolution of human replication timing in these regions. Taken together, we find significant evidence for replication timing polymorphism within both humans and chimpanzees, a substantial fraction of which appears to represent deep evolutionary processes that manifest as either conserved replication timing variation (Fig. 3*C*) or concomitant intra- and interspecies variation (Fig. 3 *D*–*F*).

Identifying intraspecific replication timing variation is particularly relevant in the context of this study since such variation can be used to map replication timing quantitative trait loci [rtQTLs; (18, 20)] which can then be tested for association with interspecies variation. Our population-level measurement of replication timing across species thus lends itself to the identification of the genetic basis of replication timing evolution.

To map rtQTLs in chimpanzees, we used fastQTL as recently described [(20); see *Methods*], controlling for the population structure and relatedness of our sample set (*SI Appendix*, Fig. S3 *C*–*F*). This unbiased genome-wide analysis identified 21 rtQTLs—a relatively small number which we ascribe to the limited sample size and relatedness of the 89 chimpanzees. To increase rtQTL discovery power, we further mapped chimpanzee rtQTLs directly in regions of chimpanzee interindividual replication timing variation and human–chimpanzee replication timing variation. This identified an additional 31 rtQTLs among the 195 chimpanzee interindividual variants and a further 33 in the 656 autosomal human–chimpanzee variant regions (*SI Appendix*, Fig. S8).

To compare replication timing polymorphisms to genetic variation in the current human LCL samples, we took advantage of the much larger number of 1,775 rtQTLs that we previously mapped in human stem cells (20) and 3,752 rtQTLs that we independently mapped in LCLs from the 1000 Genomes Project (our unpublished results). We validated 276 of the 1000 Genomes rtQTLs directly in the current human samples (of 2,793 rtQTLs for which we had all three genotypes at the top associated SNP location) and also verified that most human interindividual variants from the current study overlapped rtQTLs in the 1000 Genomes dataset (141/185, z test $P = 8.3 \times 10^{-7}$; 83 overlapped human rtQTLs validated in the current samples). Thus, rtQTLs reflect the genetic basis of replication timing variation at a substantial fraction of the sites we mapped in this study.

Since we observed a high concordance of within-species variation with between-species variation (Fig. 3 *A*–*G*), we predicted that human rtQTLs will also be associated with replication timing variation between humans and chimpanzees and could thus shed light on the genetic evolution of this form of variation. Indeed, 187 LCL and 57 iPSC human–chimpanzee variant source sites overlapped an rtQTL top associated SNP in the respective cell type significantly more than expected based on randomizations (LCL: z test $P = 3.5 \times 10^{-33}$; iPSC: z test $P = 0.004$). Since some rtQTL-associated SNPs affect replication timing at a distance, we also confirmed that source sites were enriched in rtQTL affected regions in addition to rtQTL SNPs per se (LCL: z test $P = 7.5 \times 10^{-4}$).

To test whether rtQTL sequences, at least in part, stand at the basis of replication timing evolution, we asked whether the derived allele matched the evolved replication timing state. For example, we would predict that humans carrying the ancestral allele for an rtQTL would have the ancestral replication timing (i.e., similar to chimpanzees), while humans with the derived allele would have the derived (i.e., different) replication timing state. We tested this prediction on human rtQTLs that spanned an interspecies variant source site and used the top associated rtQTL SNP and strongly linked SNPs (LD > 0.8). We found a strong enrichment of rtQTLs where the human derived allele was associated with the evolved replication timing state, while the ancestral allele was associated more closely with the chimpanzee replication timing (at least one tested SNP for 1,249/1,605 rtQTLs, 78%, permutations $P = 0.0072$; >50% of SNPs for 741/1,605, 46%, $P = 0.0012$; see examples in Fig. 3*H*). Of these rtQTLs, 227 spanned human–chimpanzee variant regions that were resolved as changes in the human lineage. In 215 of these rtQTLs (95%), the chimpanzee allele of at least one tested SNP in high LD matched the macaque allele, suggesting that the genetic association may be sustained throughout the primate lineage as well.

The same analysis for the chimpanzee rtQTLs revealed that the chimpanzee derived allele matched the evolved replication timing state for 18 of the 44 chimpanzee rtQTLs mapped in human–chimpanzee variants, suggesting that the derived chimpanzee allele was contributing to the difference in replication timing between humans and chimpanzees in these regions.

Importantly, 30 of the 44 chimpanzee rtQTL-associated regions (mapped in human–chimpanzee replication timing variants) were also shared with a human rtQTL (expected 21, z test $P = 0.005$). This was not the result of ancient polymorphisms with conserved effects on DNA replication timing as humans and chimpanzees did not share the associated rtQTL SNPs. Instead, this suggests that independent genetic contributions influence the replication timing of a given region across species, while rtQTL sharing further reflects either evolutionary pressures to maintain replication timing polymorphisms or relaxed selective constraints to fix replication timing at these loci (37, 38).

Shared Genetic Causes of Replication Timing and Gene Expression Evolution. We showed above that the evolution of DNA replication timing can be ascribed to sequence evolution,

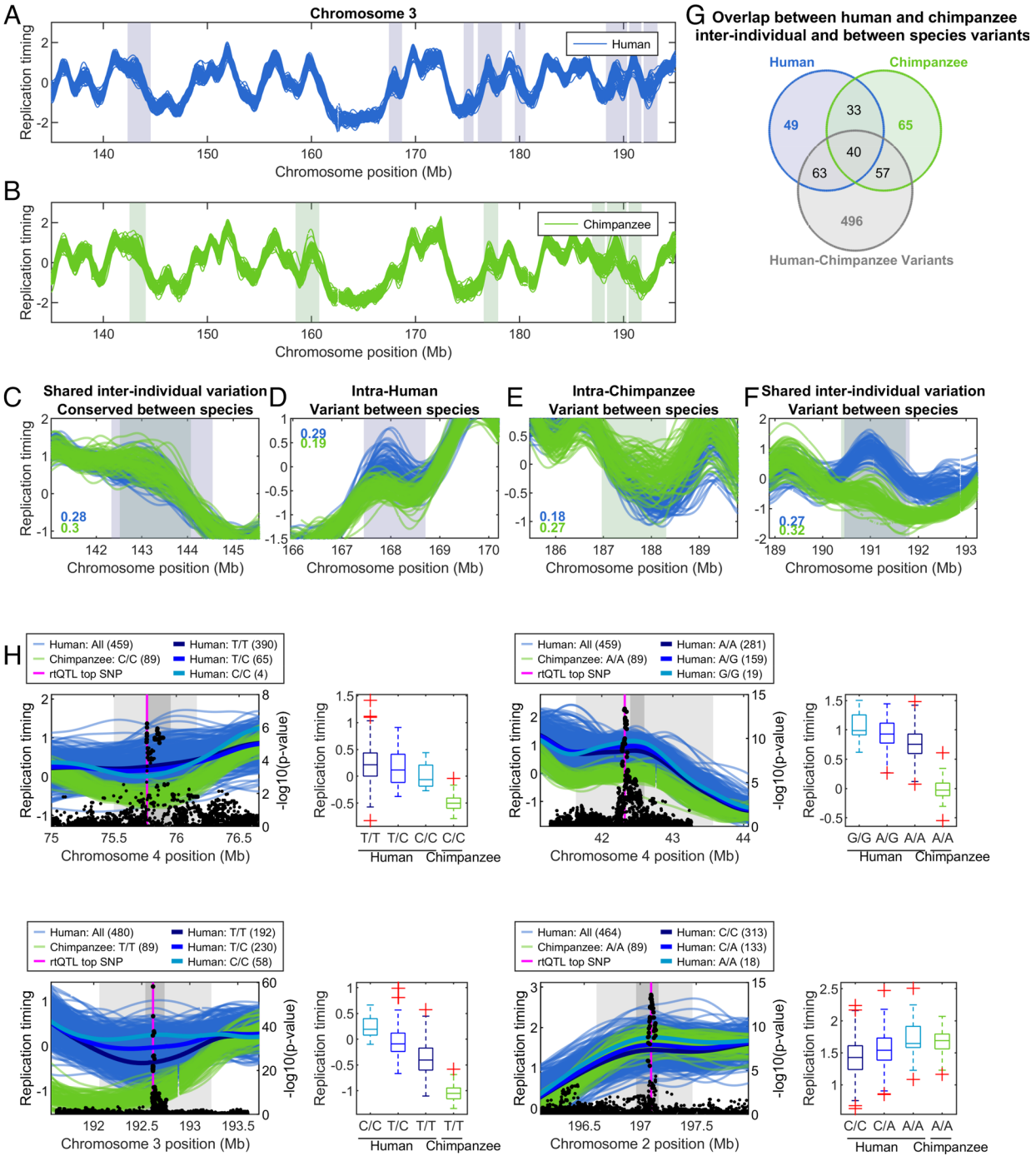

Fig. 3. Genetic variation underlying interindividual and interspecies replication timing variation. (*A* and *B*) Regions of interindividual replication timing variation in human (*A*) and chimpanzee LCLs (*B*) for a section of chromosome 3. (*C*–*F*) Examples of shared and species-specific interindividual replication timing variant regions. The numbers within each plot indicate the maximum SD of replication timing values within the variant region for human and chimpanzee (blue and green, respectively). (*G*) Overlap of intra- and inter-species replication timing variants in humans and chimpanzees. (*H*) Examples of human rtQTLs that overlap human–chimpanzee replication timing variant regions. The top rtQTL SNP (magenta) falls within, or near, the source site of replication timing variation. Human 1000 Genomes data were used for replication timing profiles and box plots. In three of the examples, replication was earlier in humans, and the chimpanzee allele at the top associated rtQTL SNP matches the late replicating human allele. The opposite direction, i.e., derived late replication in humans, is observed in the bottom right example.

while at some genomic loci, it appears that it could potentially also impact regulatory evolution. Considered jointly, and further with the sequence determinants of gene regulation, these observations could potentially reveal how gene regulation and DNA replication timing have coevolved. We previously showed that, across humans, replication timing and gene expression variation often share genetic causes (20). We thus took advantage of comprehensive mapping of gene expression QTLs (eQTLs) in LCLs by the GTEx consortium (39) and compared them to the top associated SNPs (and/or SNPs in LD > 0.8 to that top SNP) of the rtQTLs we found to be associated with replication timing variation between humans and chimpanzees. We found 488 rtQTLs (of 1,605 that overlap human–chimpanzee variants) were also significant eQTLs (q value < 0.05; 192 unique variant regions). At these eQTLs, 64% of the involved genes (194 of 301 for which expression data were available) had concordant changes in gene expression and

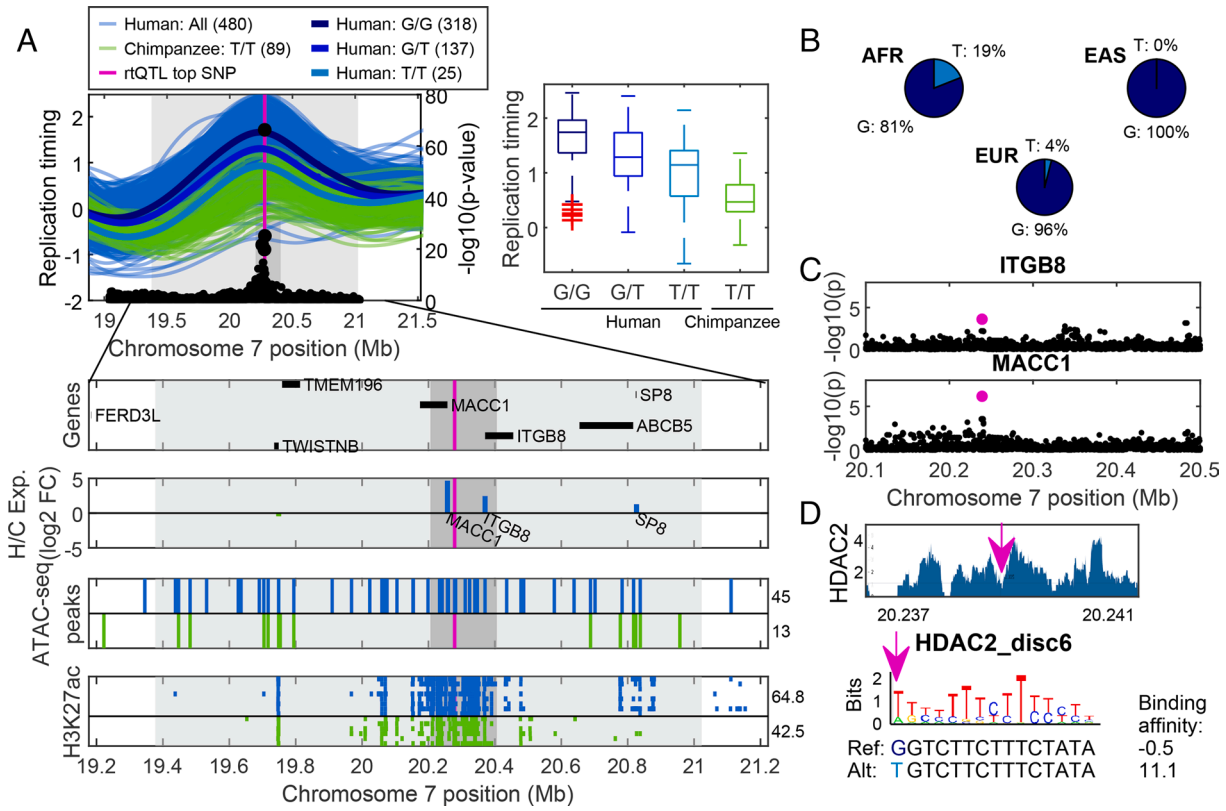

Fig. 4. A genetic variant affecting HDAC2 binding, DNA replication timing, and regional gene expression. (*A*) Chimpanzee replication timing profiles (this study) together with all replication timing profiles from the 1000 Genomes African superpopulation (AFR) and averaged replication profiles per genotype at the top associated rtQTL SNP (rs7806550). Genes, expression levels, ATAC-seq, and H3K27ac density shown below. Magenta: rs7806550 located between the genes ITGB8 and MACC1. (*B*) AFR, East Asian (EAS), and European (EUR) 1000 Genomes phase 3 allele frequencies for rs7806550. (*C*) ITGB8 and MACC1 LCL eQTLs (GTEx). Magenta: rs7806550. (*D*) An HDAC2 regulatory motif is altered by the rtQTL–eQTL top SNP (rs7806550), with the alternate allele (T) having higher binding affinity. *Top*: Encode HDAC2 ChIP-seq data for GM12878. *Bottom*: Sequence logo of HDAC2 regulatory motif that is altered by rs7806550 (magenta arrow). Binding affinities are from HaploReg (Δ affinity = 3,104-fold).

replication timing, suggesting shared genetic causes of replication timing and gene expression evolution.

A notable example was a human–chimpanzee variant region that was both an rtQTL (Fig. 4*A*) and an eQTL for two protein-coding genes (Fig. 4*C*) and one lincRNA. The rtQTL top SNP was the same as the top eQTL SNP (rs7806550), and there were no SNPs within 10 kb of the variant with LD>0.4 (Ensembl 1000 Genomes YRI LD). Among human populations, the ancestral allele frequency for rs7806550 was the highest in African populations (19%) and much lower in out-of-Africa populations (0 to 4%) (Fig. 4*B*). This suggests that the derived allele emerged in the common ancestor of humans and increased in frequency to become the major allele in modern-day humans. The region impacted by this shared rtQTL–eQTL was earlier replicating in humans than chimpanzees, and the two protein-coding genes associated with the eQTL, ITGB8 (integrin complex subunit that mediates cellular interactions) and MACC1 (regulator of hepatocyte growth factor receptor involved in cell growth and motility), were also more highly expressed in humans (Fig. 4*A*). Although rs7806550 is not known to be associated with any human phenotype [GWAS catalog; (40)], it fell within a strong LCL enhancer and was predicted to affect two transcription factor binding motifs—for GATA and for HDAC2—where the alternate allele (T; also ancestral allele) has higher binding affinity (Fig. 4*D*). HDAC2 catalyzes deacetylation of lysine residues at the N-terminal regions of core histones (H2A, H2B, H3, and H4) and we previously showed that HDAC2 binding is associated with late replicating rtQTL alleles (20). In this specific example, the ancestral allele (with higher HDAC2 binding affinity) was later replicating and matched chimpanzee replication timing. These observations can be explained if the human derived allele interrupts the HDAC2 binding site and decreases its ability to bind and deacetylate histones in the area, thus leading to greater histone acetylation, greater chromatin accessibility, and ultimately earlier replication and higher expression levels of genes in the immediate area. Interestingly, we also identified this region as variant between human and chimpanzee iPSCs (*SI Appendix*, Fig. S5 *D*, *Middle*) and previously showed it to be the location of a human iPSC rtQTL (20). Overall, this indicates that sequence changes may coordinate the concomitant evolution of replication timing and gene expression through a chromatin intermediate.

Discussion

A long-standing question in human biology is what are the genetic changes that distinguish us from other species? The significant sequence similarity between humans and our ape relatives has pointed to regulatory evolution as a likely explanation of our unique phenotypes (4, 5). However, studies of the evolution of gene expression and other epigenetic features have fallen short of fully explaining the complex adaptations in the human lineage. An understudied biological process from an evolutionary standpoint has been DNA replication timing, a fundamental genomic process that bridges genome regulation and maintenance. Previous

studies of replication timing evolution have been restricted by small sample sizes, limiting their ability to fully describe evolutionary alterations and reveal the genetic drivers and the impacts of replication timing evolution. This has also limited the understanding of the forces that drive replication timing evolution and thus understanding of its functional significance. Here, we utilized population-scale replication timing profiling of humans, chimpanzees, and rhesus macaques to identify hundreds of genomic locations that vary in replication timing within and between these species, regulatory features that covary with replication timing across species, and sequence variants associated with replication timing evolution. This study represents, to our knowledge, the most comprehensive comparison of DNA replication timing across species and one of the largest functional genomic evolutionary studies in apes; it provides a comprehensive description of a previously understudied form of human molecular evolution.

Notwithstanding local variation, the majority of the genome exhibits highly conserved replication timing both between and within species. This is unlikely even for lack of input sequence variation as we have previously shown that multiple sequence determinants, spread over areas spanning several megabases, can influence the activity of any given replication origin (20). For the replication timing variants that we do observe, most are quantitative (changes in replication origin firing times), and none can be considered of very large magnitude (e.g., >half of the S phase). Thus, it appears that DNA replication timing is largely conserved between species, consistent with previous studies (15, 16, 41). On the background of this overall conservation, we provide evidence for replication timing evolution across more than 30% of the human genome, including 123 genomic regions that have specifically evolved in the human lineage. This extent of interspecies replication timing differences illuminates the functional potential of DNA replication timing in human evolution, with sequence alterations that affect replication timing exerting a potentially broad effect on large chromosomal regions. The null assumption should be that these alterations are evolutionary neutral, which would be consistent with the observation that replication timing evolution mimics the phylogenetic tree of the same studied species. Shared variation between and within species, and shared rtQTLs [similar to previous observations of shared eQTLs; (37, 38, 42)], could also be a reflection of neutral drift and lack of local constraint. An intriguing alternative possibility in this case is the action of balancing selection, although this would be unexpected over long evolutionary timescales. The general correlations with variation in gene regulation and potentially with mutation rates (see further below) point to the possibility that a subset of replication timing-evolved regions could carry important functional implications for human evolution.

It is notable that we did not identify any obvious fixed difference in replication timing between humans and chimpanzees, despite the numerous fixed sequence changes between the species. This may in part be due to our limited ability to call fixed differences in replication timing: Most variants are quantitative changes in origin firing time and are thus difficult to categorize as fixed or polymorphic changes. Even so, it is interesting to consider whether the unexpected lack of fixed differences could indicate selection against fixed changes in replication origin activity, which can be expected to impact large chromosomal domains. Another possibility is ongoing evolution of the replication timing program, potentially pointing to the neutrality of such changes. We note, however, that the evolutionary divergence time of humans and chimpanzees makes ongoing selection unexpected, while balancing selection is itself uncommon. Another important point to consider is that multiple independent sequence changes are often required for a change in replication timing (20). This provides a plausible explanation for how fixed changes between the species could translate to polymorphic changes in replication timing without ruling out selection on specific changes that alter replication timing. Overall, our results point to intriguing evolutionary dynamics of the replication timing program, which cannot be immediately explained by sequence evolution.

Studying a large number of individuals provided the unique ability to detect replication timing variation concomitantly between and within species. We observed an extensive overlap of variation within and between species, pointing to deep and ongoing evolutionary processes impacting replication timing. This overlap also highlights the value of large sample sizes in evolutionary studies of replication timing since variation between species is likely obscured in studies using smaller sample sizes. Furthermore, identifying and comparing variation within and between species enabled us, for the first time (to our knowledge), to reveal some of the genetic determinants of human replication timing evolution. We did this using an rtQTL mapping approach, which we applied to each species separately and then combined by considering co-occurring interspecies replication timing variation and linking derived rtQTL alleles to derived replication timing states. We anticipated that a much larger fraction of replication timing variants is determined by sequence evolution and could have been revealed with a larger sample size and hence greater power to detect rtQTLs.

Identification of genetic determinants of replication timing evolution provides a means for revealing mechanisms of replication timing control and the population genetic and evolutionary dynamics of replication timing. A notable example we highlighted pertains to the role of histone acetylation. HDAC binding has been described previously as a repressor of replication timing (20, 43), likely by promoting a more repressive chromatin state. Here, we showed that a sequence polymorphism impacting an HDAC2 binding site within a human LCL enhancer has likely led to variation in both replication timing and gene expression within humans and between humans and chimpanzees. It is notable that the derived allele is present at a relatively low (19%) frequency in African individuals and has risen to near fixation in out-of-Africa populations.

More generally, and consistent with previous studies, we observed correlated covariation of DNA replication timing, chromatin accessibility, and gene expression. A notable advantage of our study, however, is the ability to interrogate these relationships across hundreds of genomic regions harboring natural variation in DNA replication timing. While previous studies typically described these correlations as indicative of unidirectional causality relationships [in either direction; (14, 44–48)], our study suggests a more complex picture of replication timing and gene expression covarying and potentially affecting each other (with chromatin structure being a likely intermediary) in a cell-type–specific and locus- and context-specific manner. While it is intriguing to consider a potential influence of replication timing on the establishment of chromatin structure (and gene expression patterns), it is important to keep in mind that there were many instances of decorrelation between these processes. Thus, these processes are fundamentally regulated independently of each other.

Another important functional aspect of replication timing is its influence on the mutational landscape and therefore on local sequence evolution. Beyond validating the correlation between replication timing and rates of mutation and sequence variation, we found that sequence divergence between humans and

chimpanzees was elevated specifically in replication timing variant regions (in iPSCs in particular). This suggests that evolutionary changes in replication timing could potentially alter local mutation rates and patterns. Another way to explain this relationship would be if sequence alterations impacting replication timing are more likely to occur in genomic regions with increased divergence. However, we favor the former interpretation since replication timing has been shown to be one of the strongest (if not the single strongest) correlates with megabase-scale variation in genomic divergence rates (10, 49). An impact of replication timing variation on mutation rates would also be consistent with the observation that protein-coding genes (especially loss-of-function–intolerant genes) were generally depleted in the same variant regions; thus, replication timing alterations in regions with highly conserved genomic function may be unfavorable, possibly due to the dual relationship of replication timing with genome regulation and genome maintenance.

More comprehensive mutational data would be required in order to test with sufficient statistical power how replication timing affects sequence evolution. The ability to obtain additional replication timing and mutation data for chimpanzees is, however, notably limited by the scarcity and regulatory limitations of using chimpanzee material. An alternative is to study replication timing variation within the more limited evolutionary timescale of human populations or the broader timescale of more diverged mammalian species such as rodents compared to primates.

Another critical observation is that replication timing evolution is predominantly cell-type specific. iPSCs and LCLs had a very little overlap of interspecies replication timing variants. By inference, other cell types can be expected to show replication timing evolution at yet other genomic locations, and therefore, the full functional impact of replication timing evolution would only be possible to evaluate in a larger number of cell types.

Overall, our findings highlight the extent and potential importance of replication timing evolution as both a driver and consequence of sequence and regulatory evolution at certain regions of the genome. DNA replication timing may mark an important yet previously underconsidered form of human genome evolution with potential interfaces with past and/or future functional evolution. Several important questions remain, including the role of replication timing in human evolution, the evolutionary forces shaping replication timing evolution, the causal direction of replication timing with chromatin and gene expression alterations, and the influence of replication timing evolution on the mutational landscape. As such, it would be highly informative to incorporate replication timing in future studies of sequence and epigenetic evolution.

The extent of differences in replication timing between humans and chimpanzees was previously unknown. Our finding of several hundred regions of variation, and the precise genetic mapping of many, provides an opportunity to edit (e.g., in cell lines or mice) rtQTL sequences that contribute to human replication timing evolution and thus elucidate the genomic, cellular, and phenotypic impacts of replication timing.

Methods

Genomic DNA for 90 chimpanzee LCLs, 88 human LCLs, 23 rhesus macaque LCLs, seven human iPSCs, and seven chimpanzee iPSCs were sequenced to approximately 20× coverage using the Illumina HiSeq X Ten with 2 × 150 paired-end reads. Sequence data were aligned to their respective reference genomes (hg19, panTro6, and rheMac10) using BWA–MEM, and replication timing profiles were generated with TIGER as previously described (19). To compare replication timing profiles between species, we used the UCSC genome browser liftOver tool with reciprocal best mapping chains to convert chimpanzee and rhesus coordinates to human coordinates. Replication timing variants between humans and chimpanzees were identified with ANOVA tests in 200-kb sliding windows across the genome. Evolutionary changes were inferred using rhesus macaque as an outgroup by calculation of pairwise Euclidean distances. The association of replication timing with gene density, gene expression, chromatin accessibility, and sequence variation was performed genome wide by binning replication timing windows into 30 equally portioned bins and counting the number of genomic features within each bin. Human and chimpanzee chromatin peaks (ATAC-seq and ChIP-seq for H3K27ac, H3K27me3, H3K4me1, and H3K4me3) were also counted within each replication timing variant region, and the $\log_2$ fold change in peak density was calculated to associate changes in replication timing to changes in chromatin across species. This was similarly applied to gene expression changes. Interindividual replication timing variants were identified by calculating replication timing SD across the genome followed by pairwise *t* tests. Chimpanzee SNPs and indels were called with GATK (50) and used in the rtQTL analysis. Chimpanzee rtQTLs were mapped genome wide in human–chimpanzee variants and within chimpanzee replication timing variants using fastQTL as in ref. 20. To identify shared genetic causes of replication timing and expression variation, we located GTEx LCL eQTLs that shared a top associated rtQTL SNP.

Data, Materials, and Software Availability. Raw chimpanzee and rhesus macaque whole-genome sequencing data are available under SRA accession PRJNA856315 (51), and human data are available under dbGaP accession phs002597 (52). Three human LCL and six human iPSC samples were not consented for release of raw genomic sequence data. Processed replication timing data are available for all samples (thekorenlab.org/data) (53), and raw read locations are available from the authors by request.

ACKNOWLEDGMENTS. We thank Ana Rita Rebelo, Bronte Zhang, Lauren Mei, Sean Kim, and Tiffany Ge for technical assistance and members of our laboratory for critical reading of the manuscript. We thank Yoav Gilad for sharing human and chimpanzee iPSC and rhesus macaque LCL samples and for fruitful discussions. Chimpanzee LCL samples were sourced from the Yerkes Center (grant no. P51OD011132); DONSON human LCLs were a gift from Grant Stewart. This work was supported by the NSF (MCB-1921341 to Amnon Koren). A.N.B. was partially supported by the Cornell University Center for Vertebrate Genomics. A.D. is a Hunter R. Rawlings III Cornell Presidential Research Scholars.

Author affiliations: [a]Department of Molecular Biology and Genetics, Cornell University, Ithaca, NY 14853; and [b]Bioinformatics Facility, Institute of Biotechnology, Cornell University, Ithaca, NY 14853

1. M. V. Suntsova, A. A. Buzdin, Differences between human and chimpanzee genomes and their implications in gene expression, protein functions and biochemical properties of the two species. *BMC Genomics* **21**, 1–12 (2020).
2. I. G. Romero, I. Ruvinsky, Y. Gilad, Comparative studies of gene expression and the evolution of gene regulation. *Nat. Rev. Genetics* **13**, 505–516 (2012).
3. Z. N. Kronenberg *et al.*, High-resolution comparative analysis of great ape genomes. *Science* **360**, eaar6343 (2018).
4. H. B. Fraser, Gene expression drives local adaptation in humans. *Genome Res.* **23**, 1089–1096 (2013).
5. M.-C. King, A. C. Wilson, Evolution at two levels in humans and chimpanzees: Their macromolecules are so alike that regulatory mutations may account for their biological differences. *Science* **188**, 107–116 (1975).
6. D. Gokhman *et al.*, Human–chimpanzee fused cells reveal cis-regulatory divergence underlying skeletal evolution. *Nat. Genet.* **53**, 467–476 (2021).
7. R. M. Agoglia *et al.*, Primate cell fusion disentangles gene regulatory divergence in neurodevelopment. *Nature* **592**, 421–427 (2021).
8. D. J. Massey, A. Koren, High-throughput analysis of single human cells reveals the complex nature of DNA replication timing control. *Nat. Commun.* **13**, 1–15 (2022).
9. N. Rhind, D. M. Gilbert, DNA replication timing. *Cold Spring Harb. Perspect. Biol.* **5**, a010132 (2013).
10. J. A. Stamatoyannopoulos *et al.*, Human mutation rate associated with DNA replication timing. *Nat. Genet.* **41**, 393–395 (2009).
11. A. Koren *et al.*, Differential relationship of DNA replication timing to different forms of human mutation and variation. *Am. J. Hum. Genet.* **91**, 1033–1040 (2012).
12. L. C. Francioli *et al.*, Genome-wide patterns and properties of de novo mutations in humans. *Nat. Genet.* **47**, 822–826 (2015).
13. N. Agier *et al.*, The evolution of the temporal program of genome replication. *Nat. Commun.* **9**, 1–12 (2018).

14. C. A. Müller, C. A. Nieduszynski, DNA replication timing influences gene expression level. *J. Cell Biol.* **216**, 1907–1914 (2017).
15. E. Yaffe *et al.*, Comparative analysis of DNA replication timing reveals conserved large-scale chromosomal architecture. *PLoS Genet.* **6**, e1001011 (2010).
16. T. Ryba *et al.*, Evolutionarily conserved replication timing profiles predict long-range chromatin interactions and distinguish closely related cell types. *Genome Res.* **20**, 761–770 (2010).
17. Y. Yang, Continuous-trait probabilistic model for comparing multi-species functional genomic data. *Cell Syst.* **7**, 208–218.e11 (2018).
18. A. Koren *et al.*, Genetic variation in human DNA replication timing. *Cell* **159**, 1015–1026 (2014).
19. A. Koren, D. J. Massey, A. N. Bracci, TIGER: Inferring DNA replication timing from whole-genome sequence data. *Bioinformatics* **37**, 4001–4005 (2021).
20. Q. Ding *et al.*, The genetic architecture of DNA replication timing in human pluripotent stem cells. *Nat. Commun.* **12**, 1–18 (2021).
21. R. S. Hansen *et al.*, Sequencing newly replicated DNA reveals widespread plasticity in human replication timing. *Proc. Natl. Acad. Sci. U.S.A.* **107**, 139–144 (2010).
22. A. Koren, S. A. McCarroll, Random replication of the inactive X chromosome. *Genome Res.* **24**, 64–69 (2014).
23. D. C. Soto *et al.*, Identification of structural variation in chimpanzees using optical mapping and nanopore sequencing. *Genes* **11**, 276 (2020).
24. I. G. Romero *et al.*, A panel of induced pluripotent stem cells from chimpanzees: A resource for comparative functional genomics. *Elife* **4**, e07103 (2015).
25. X. Zhou *et al.*, Epigenetic modifications are associated with inter-species gene expression variation in primates. *Genome Biol.* **15**, 547 (2014).
26. Z. Khan *et al.*, Primate transcript and protein expression levels evolve under compensatory selection pressures. *Science* **342**, 1100–1104 (2013).
27. R. García-Pérez *et al.*, Epigenomic profiling of primate lymphoblastoid cell lines reveals the evolutionary patterns of epigenetic activities in gene regulatory architectures. *Nat. Commun.* **12**, 1–17 (2021).
28. F. Tajima, Simple methods for testing the molecular evolutionary clock hypothesis. *Genetics* **135**, 599–607 (1993).
29. E. J. Vallender, B. T. Lahn, Positive selection on the human genome. *Hum. Mol. Genet.* **13**, R245–R254 (2004).
30. M. Gayà-Vidal, M. M. Albà, Uncovering adaptive evolution in the human lineage. *BMC Genomics* **15**, 599 (2014).
31. S. L. Sawyer *et al.*, Ancient adaptive evolution of the primate antiviral DNA-editing enzyme APOBEC3G. *PLoS Biol.* **2**, e275 (2004).
32. Y. Ding *et al.*, Antitumor activity of IFIX, a novel interferon-inducible HIN-200 gene, in breast cancer. *Oncogene* **23**, 4556–4566 (2004).
33. M. J. Hubisz, K. S. Pollard, Exploring the genesis and functions of Human Accelerated Regions sheds light on their role in human evolution. *Curr. Opin. Genet. Dev.* **29**, 15–21 (2014).
34. S. Peyrégne *et al.*, Detecting ancient positive selection in humans using extended lineage sorting. *Genome Res.* **27**, 1563–1572 (2017).
35. K. J. Karczewski *et al.*, The mutational constraint spectrum quantified from variation in 141,456 humans. *Nature* **581**, 434–443 (2020).
36. G. Bejerano *et al.*, Ultraconserved elements in the human genome. *Science* **304**, 1321–1325 (2004).
37. B. J. Fair *et al.*, Gene expression variability in human and chimpanzee populations share common determinants. *Elife* **9**, e59929 (2020).
38. J. Tung *et al.*, The genetic architecture of gene expression levels in wild baboons. *ELife* **4**, e04729 (2015).
39. J. Lonsdale *et al.*, The genotype-tissue expression (GTEx) project. *Nat. Genet.* **45**, 580–585 (2013).
40. A. Buniello *et al.*, The NHGRI-EBI GWAS catalog of published genome-wide association studies, targeted arrays and summary statistics 2019. *Nucleic Acids Res.* **47**, D1005–D1012 (2019).
41. C. A. Müller, C. A. Nieduszynski, Conservation of replication timing reveals global and local regulation of replication origin activity. *Genome Res.* **22**, 1953–1962 (2012).
42. A. J. Jasinska *et al.*, Genetic variation and gene expression across multiple tissues and developmental stages in a nonhuman primate. *Nat. Genet.* **49**, 1714–1721 (2017).
43. A. Goren *et al.*, DNA replication timing of the human β-globin domain is controlled by histone modification at the origin. *Genes Dev.* **22**, 1319–1324 (2008).
44. K. N. Klein *et al.*, Replication timing maintains the global epigenetic state in human cells. *Science* **372**, 371–378 (2021).
45. J. C. Rivera-Mulia *et al.*, Dynamic changes in replication timing and gene expression during lineage specification of human pluripotent stem cells. *Genome Res.* **25**, 1091–1103 (2015).
46. M. Blin *et al.*, Transcription-dependent regulation of replication dynamics modulates genome stability. *Nat. Struct. Mol. Biol.* **26**, 58–66 (2019).
47. J. Zhang *et al.*, Establishment of transcriptional competence in early and late S phase. *Nature* **420**, 198–202 (2002).
48. C. Marchal, J. Sima, D. M. Gilbert, Control of DNA replication timing in the 3D genome. *Nat. Rev. Mol. Cell Biol.* **20**, 721–737 (2019).
49. C.-L. Chen *et al.*, Impact of replication timing on non-CpG and CpG substitution rates in mammalian genomes. *Genome Res.* **20**, 447–457 (2010).
50. G. A. Van der Auwera *et al.*, From FastQ data to high-confidence variant calls: The genome analysis toolkit best practices pipeline. *Curr. Protoc. Bioinform.* **43**, 11.10.1–11.10.33 (2013).
51. A. N. Bracci, A. Koren, Whole-genome sequencing of chimpanzee and rhesus macaque cell lines. NCBI SRA. https://www.ncbi.nlm.nih.gov/sra/PRJNA856315. Deposited 6 July 2022.
52. M. Caballero, A. N. Bracci, A. Koren, DNA replication timing alterations in genetic diseases. NCBI SRA. https://www.ncbi.nlm.nih.gov/gap/advanced_search/?TERM=phs002597. Deposited 6 June 2022.
53. A. N. Bracci, A. Koren, DNA replication timing in humans, chimpanzees and rhesus macaques. The Koren Lab. https://www.thekorenlab.org/data. Deposited 6 June 2022.